# CS70 @ UC Berkeley, Spring 2026

## Lecture 21

## Random Variables II:
## Variance and Covariance

April 9, 2026

# Get the Largest Number (Lecture 20)

Consider a deck of $N$ cards each with a number written on one side, facing down. **Assume:**

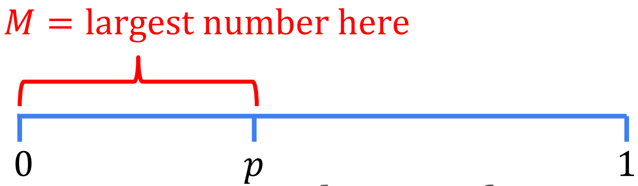
- $N$ is large and all numbers are distinct.
- The deck is well shuffled.

**Goal:** Get the largest number.

**Rules:**

1. Reveal one card at a time starting from top.
2. STOP at the current card or reveal the next card.
3. If you pass on a card that has been revealed, you can't choose to it later.

**Strategy:** Reveal a certain proportion, say $p$, of the cards and record the largest number, denoted $M$, you have seen. Then, STOP when you see a number larger than $M$.




**Question:** What should $p$ be to maximize your chance of winning?

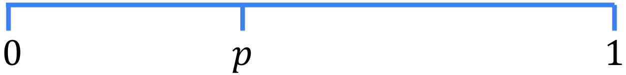
# Get the Largest Number (Lecture 20)

Success = get the largest number. $\mathbb{P}(\text{success})$?

Denote the sorted numbers as $X_1 > X_2 > \cdots > X_N$

**Case 1:** $M = X_1 \;\Rightarrow\;$ fail

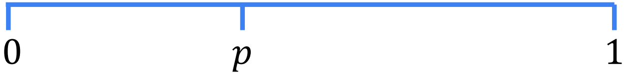
$\mathbb{P}(M = X_1) = p$




$\mathbb{P}(\text{success}|M = X_1) = 0$

**Case 2:** $M = X_2 \;\Rightarrow\; X_1$ is in $(p, 1]$

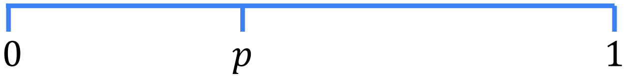
$\mathbb{P}(M = X_2) \approx p(1-p)$




$\mathbb{P}(\text{success}|M = X_2) = 1$

**Case 3:** $M = X_3 \;\Rightarrow\; X_1, X_2$ are in $(p, 1]$

$\mathbb{P}(M = X_3) \approx p(1-p)^2$




$\mathbb{P}(\text{success}|M = X_3) = \mathbb{P}(X_1 \text{ appears before } X_2 \text{ in } (p, 1]) = \frac{1}{2}$

**General Case:** $\mathbb{P}(M = X_{k+1}) \approx p(1-p)^k$ Good approximation for $k \ll N$

$\mathbb{P}(\text{success}|M = X_{k+1}) = \frac{1}{k}$

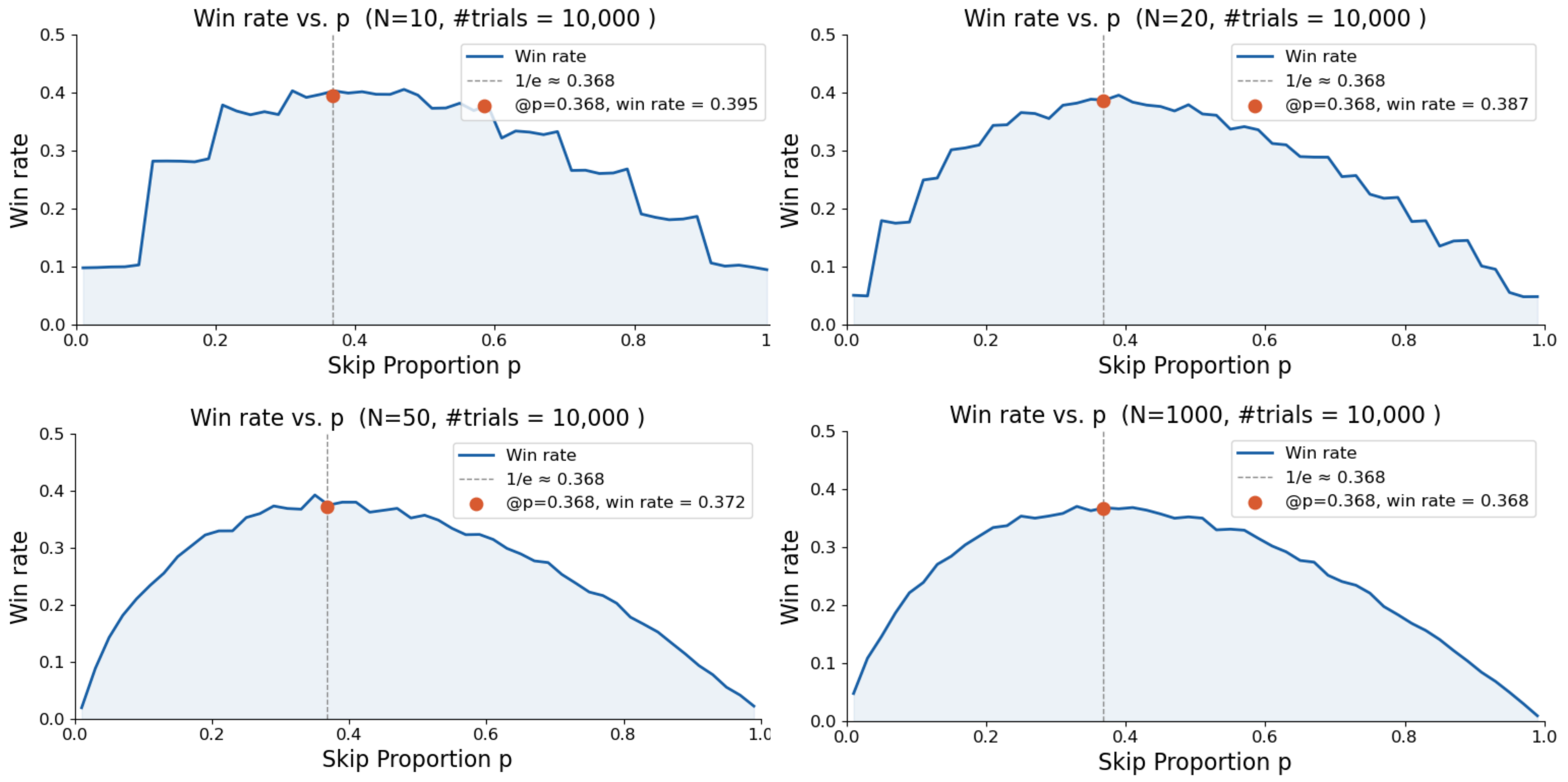
$$\mathbb{P}(\text{success}) = \sum_{i=1}^{(1-p)N} \mathbb{P}(\text{success}|M = X_i)\mathbb{P}(M = X_i)$$

decays exponentially fast as $k$ increases

$$\approx \sum_{k=1}^{(1-p)N} \frac{1}{k} p(1-p)^k \approx -p \ln p$$

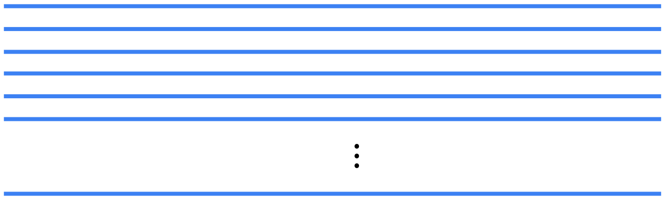
this is maximized when $p = \frac{1}{e} \approx 0.368$

# Get the Largest Number (Simulation vs. Theory)



# Average Number of Loops

$n$ identical strings

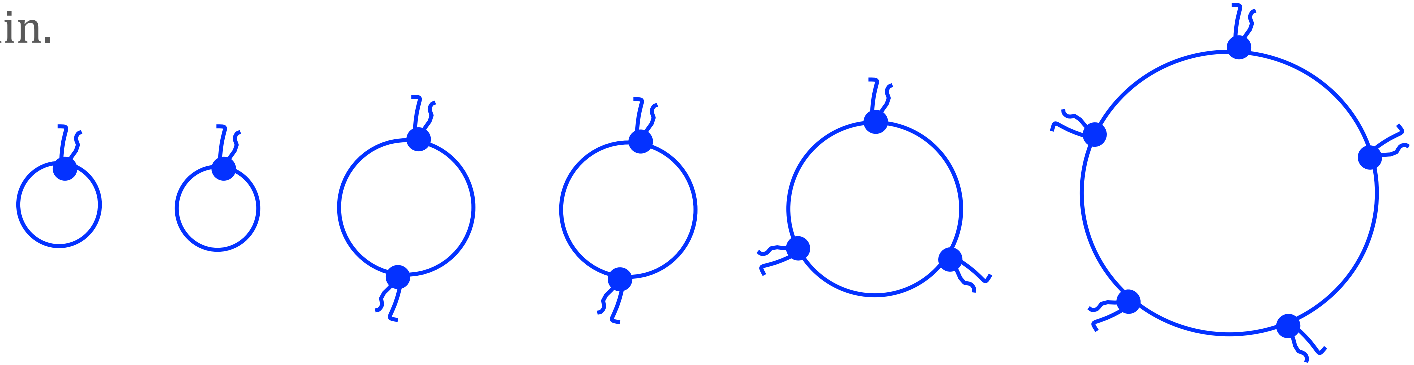


Fold them in halves and put the folded ends in a bag

$2n$ ends

Suppose you tie two free ends uniformly at random and continue until no more free ends remain.

Example outcome for $n = 14$ :



6 loops formed

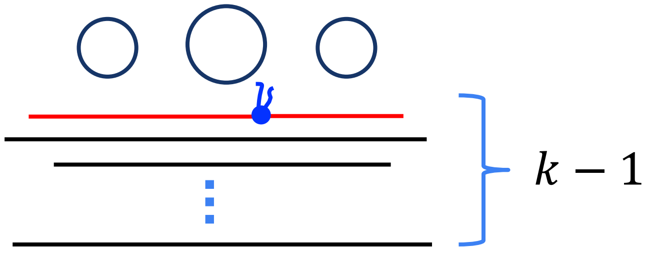
**Question:** What is the **expected** number of loops formed?

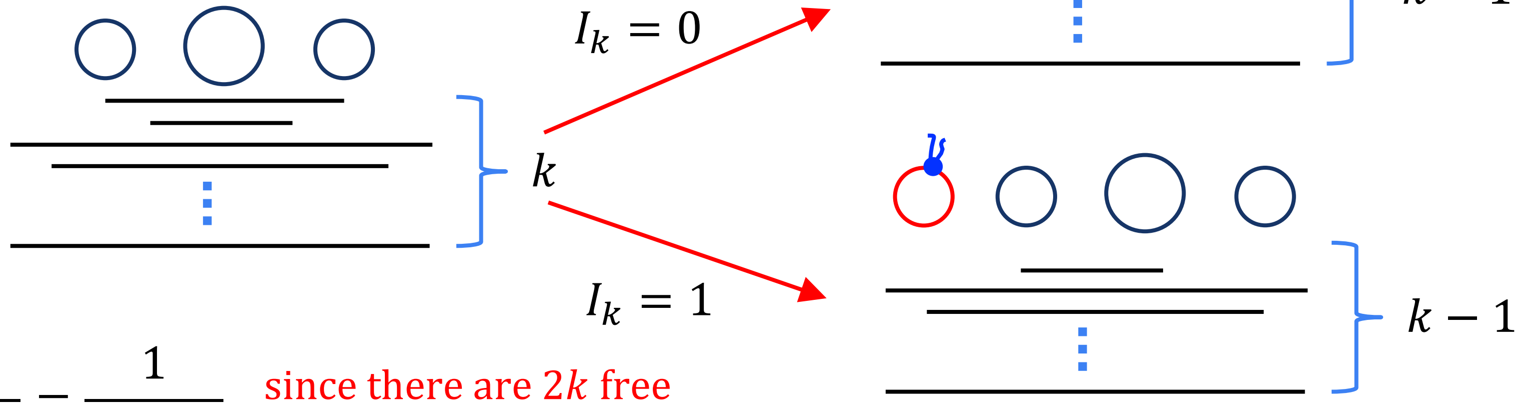
# Average Number of Loops

$L\ =$ Number of Loops formed. $\mathbb{E}[L]$?

Suppose currently there are $k$ open strings

$L = I_n + I_{n-1} + \cdots + I_1$ where $I_k =$ Indicator for the event that a loop is formed when currently there are $k$ open strings.




$$\mathbb{P}(I_k = 1) = \frac{k}{\binom{2k}{2}} = \frac{1}{2k-1}$$

since there are $2k$ free ends and $k$ open strings

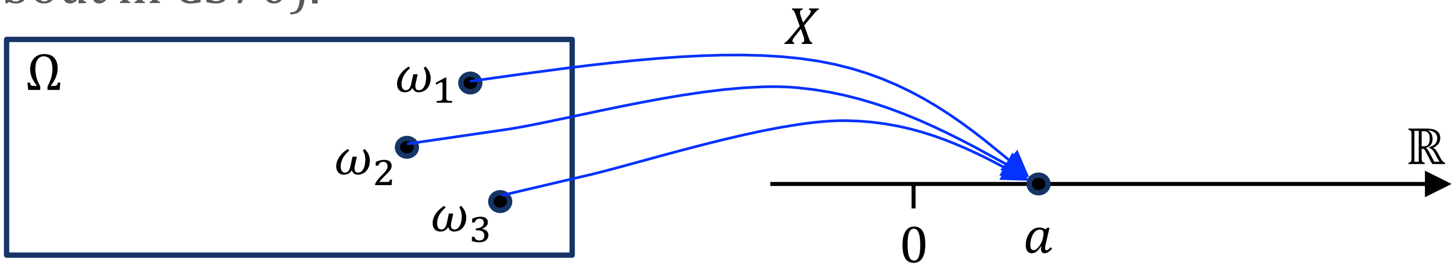
$$\mathbb{E}[L] = \mathbb{E}[I_n + I_{n-1} + \cdots + I_1] = \mathbb{E}[I_n] + \mathbb{E}[I_{n-1}] + \cdots + \mathbb{E}[I_1] = \sum_{k=1}^{n} \mathbb{P}[I_k = 1] = 1 + \frac{1}{3} + \cdots + \frac{1}{2n-1}$$

which is $\sim \frac{1}{2}\ln(n)$ for large $n$

# Review

- A random variable is a function $X: \Omega \to \mathbb{R}^n$ (we will mostly deal with the $n = 1$ case) satisfying some technical conditions (which you will not need to worry about in CS70).

Probability measure $\mathbb{P}$ is defined on the set of subsets of $\Omega$




- Let $\mathcal{A} = \{a \in \mathbb{R} \mid X(\omega) = a \text{ for some } \omega \in \Omega\}$ denote the range of $X$.
- For a given $a \in \mathcal{A}$, the event "$X = a$" is defined as the pre-image

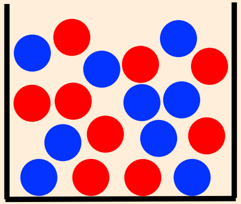
$$X^{-1}(a) = \{\omega \in \Omega \mid X(\omega) = a\} \subseteq \Omega.$$

**Example:** Toss a fair coin twice. $X(\omega) =$ #number of **Heads** in $\omega \in \Omega$.
"$X = 2$" $= X^{-1}(2) = \{(H, H)\}$
"$X = 1$" $= X^{-1}(1) = \{(H, T), (T, H)\}$
"$X = 0$" $= X^{-1}(0) = \{(T, T)\}$
$\mathbb{P}(X = 2) = \mathbb{P}(X = 0) = ¼$, and $\mathbb{P}(X = 1) = ½$.

# Review

Sample once. Success probability $p = B/N$.

**Example (Bernoulli):** $I \sim \text{Bernoulli}(p)$, for $0 < p < 1$.



$$\mathbb{P}(I = a) = \begin{cases} p, & \text{if } a = 1, \\ 1 - p, & \text{if } a = 0. \end{cases}$$

Sample $n$ times **with** replacement.

**Example (Binomial):** $X \sim \text{Binomial}(n, p)$, where $n \in \mathbb{Z}_+$ and $0 < p < 1$.

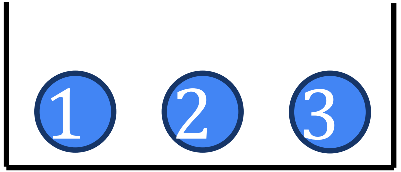
For $k = 0, \dots, n$, $\quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Sample $n$ times **without** replacement.

**Example (Hypergeometric):** $Y \sim \text{Hypergeometric}(N, B, n)$

For $\max(0, n + B - N) \le k \le \min(n, B)$, $\quad \mathbb{P}(Y = k) = \dfrac{\binom{B}{k}\binom{N-B}{n-k}}{\binom{N}{n}}$.

# Review




Sample 2 balls at random

$X =$ label of the first ball

$Y =$ label of the second ball

## Case 1: Sample with replacement

| | $X=1$ | $X=2$ | $X=3$ | Row sums |
|---|---|---|---|---|
| $Y=1$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{3}$ |
| $Y=2$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{3}$ |
| $Y=3$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{3}$ |
| Column sums | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | |

X, Y are independent and identically distributed (i.i.d.)

## Case 2: Sample without replacement

| | $X=1$ | $X=2$ | $X=3$ | Row sums |
|---|---|---|---|---|
| $Y=1$ | 0 | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{3}$ |
| $Y=2$ | $\frac{1}{6}$ | 0 | $\frac{1}{6}$ | $\frac{1}{3}$ |
| $Y=3$ | $\frac{1}{6}$ | $\frac{1}{6}$ | 0 | $\frac{1}{3}$ |
| Column sums | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | |

$X, Y$ are **not** independent, but identically distributed

# Review

- Indicator Random Variables
- For $i = 1, \dots, n$, let

$$I_i(\omega) = \begin{cases} 1, & \text{if the } i\text{th trial in } \omega \text{ is success,} \\ 0, & \text{otherwise.} \end{cases}$$

- For $X \sim \text{Binomial}(n, p)$, $X = I_1 + \cdots + I_n$, where $I_1, \dots, I_n$ are mutually independent Bernoulli($p$) random variables.

- For $Y \sim \text{Hypergeometric}(N, B, n)$, $Y = I_1 + \cdots + I_n$, where $I_1, \dots, I_n$ are exchangeable Bernoulli($p$) random variables with $p = B/N$.

**Definition (Exchangeability):** A collection of random variables $X_1, \dots, X_n$ with common range $\mathcal{A}$ is said to be exchangeable if $\left(X_{\pi(1)}, \dots, X_{\pi(n)}\right)$ has the same joint distribution as $(X_1, \dots, X_n)$ for every permutation $\pi$ of $\{1, \dots, n\}$.

# Review

Arithmetic average of 1, 1, 1, 3, 3, 4

Value Frequency

$$= \frac{1+1+1+3+3+4}{6} = \frac{1\times 3 + 3\times 2 + 4\times 1}{6} = 1\times\frac{3}{6} + 3\times\frac{2}{6} + 4\times\frac{1}{6}$$

**Definition (Expectation).** The expectation of a random variable $X$ with range $\mathcal{A}$ is defined as

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega) = \sum_{a\in\mathcal{A}} a\,\mathbb{P}(X=a),$$

provided that $\sum_{a\in\mathcal{A}} |a|\,\mathbb{P}(X=a) < \infty$ (absolutely convergent).

**Theorem (Linearity of Expectation):** For any two random variables $X$ and $Y$ on the same probability space and fixed **constants** $\alpha$ and $\beta$,

$$\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y].$$

(Note: this holds irrespective of whether $X$ and $Y$ are independent.)

# Examples

**Example (Bernoulli):** $I \sim \text{Bernoulli}(p)$, for $0 < p < 1$.

$$\mathbb{P}(I = a) = \begin{cases} p, & \text{if } a = 1, \\ 1 - p, & \text{if } a = 0. \end{cases}$$

$$\mathbb{E}[I] = 1 \cdot \mathbb{P}(I = 1) + 0 \cdot \mathbb{P}(I = 0) = p$$

**Example (Binomial):** $X \sim \text{Binomial}(n, p)$, where $n \in \mathbb{Z}_+$ and $0 < p < 1$.
$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, for $k = 0, \dots, n$.

$$\mathbb{E}[X] = \sum_{k=0}^{n} k\, \mathbb{P}(X = k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

Cumbersome calculation

$$\mathbb{E}[X] = \mathbb{E}[I_1 + \cdots + I_n] = \mathbb{E}[I_1] + \cdots + \mathbb{E}[I_n] = np$$

Much easier

Linearity of expectation

# Examples

**Example (Hypergeometric):** $Y \sim \text{Hypergeometric}(N, B, n)$

$$\mathbb{P}[Y = k] = \frac{\binom{B}{k}\binom{N-B}{n-k}}{\binom{N}{n}}.$$

$$\mathbb{E}[Y] = \sum_{k=0}^{n} k\,\mathbb{P}(Y = k) = \sum_{k=0}^{n} k\,\frac{\binom{B}{k}\binom{N-B}{n-k}}{\binom{N}{n}}$$

Cumbersome calculation

$$\mathbb{E}[Y] = \mathbb{E}[I_1 + \cdots + I_n] = \mathbb{E}[I_1] + \cdots + \mathbb{E}[I_n] = n\mathbb{E}[I_1] = n\frac{B}{N}$$

Much easier

Linearity of expectation

Exchangeability

# Examples

**Example (Fixed points, Lecture 15):**
$\Omega =$ set of all permutations of $\{1, \dots, n\}$.
Suppose $\mathbb{P}[\omega] = \frac{1}{|\Omega|} = \frac{1}{n!}$ for all $\omega \in \Omega$
Let $X(\omega) =$ denote the number of fixed points of $\omega \in \Omega$.
What is $\mathbb{E}[X]$?

$$\begin{array}{rccccccc} i: & 1 & 2 & \boxed{3} & 4 & \boxed{5} & \dots & n \\ \text{Permutation } \pi_i: & 7 & 1 & \boxed{3} & 8 & \boxed{5} & \dots & 2 \end{array}$$

Fixed points

Let $I_i(\omega) = \begin{cases} 1, & \text{if } i \text{ is a fixed point of } \omega \in \Omega, \\ 0, & \text{otherwise.} \end{cases}$

By linearity of expectation

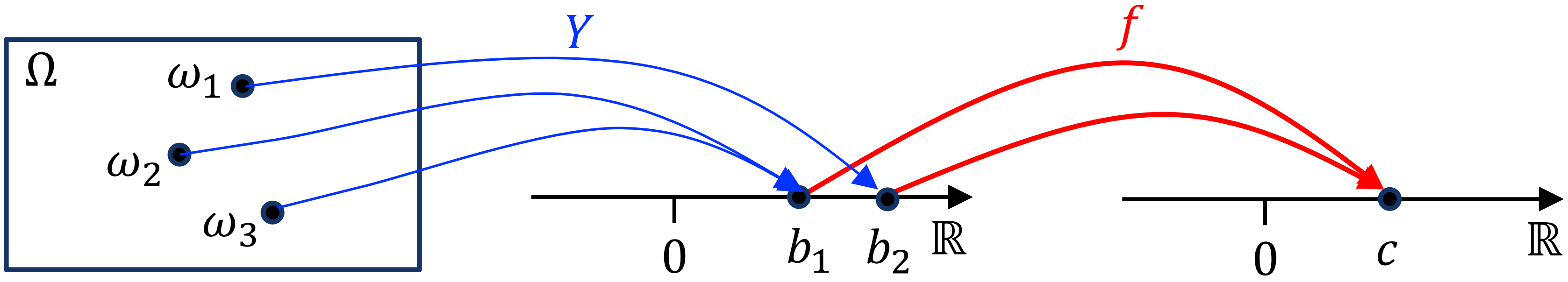
Then, $X = I_1 + \cdots + I_n$ and $\mathbb{E}[X] = \mathbb{E}[I_1 + \cdots + I_n] = \mathbb{E}[I_1] + \cdots + \mathbb{E}[I_n]$

These indicators are NOT independent

$\mathbb{P}[I_i = 1] = \dfrac{(n-1)!}{n!} = \dfrac{1}{n}$ and $\mathbb{E}[I_i] = \dfrac{1}{n}$ for all $i = 1, \dots, n$.

Therefore, $\mathbb{E}[X] = 1$. Does not depend on $n$.

# Functions of Random Variables




- Suppose $Y: \Omega \to \mathbb{R}$ is a random variable.
- Suppose the range (or image) of $Y$ is $Y(\Omega) = \mathcal{B}$.
- $f: \mathcal{B} \to \mathbb{R}$, a function satisfying a certain technical condition which we will not need to worry about in CS70.
- Then, $f(Y)$: $\Omega \to \mathbb{R}$ is also a random variable with range $\mathcal{C} = f(\mathcal{B})$ .
- $\mathbb{P}(f(Y) = c) = \sum_{b \in \mathcal{B}: f(b)=c} \mathbb{P}(Y = b)$
- $\mathbb{E}[f(Y)] = \sum_{c \in \mathcal{C}} c\, \mathbb{P}(f(Y) = c) = \sum_{c \in \mathcal{C}} c \sum_{b \in \mathcal{B}: f(b)=c} \mathbb{P}(Y = b)$

$$= \sum_{c \in \mathcal{C}} \sum_{b \in \mathcal{B}: f(b)=c} f(b)\mathbb{P}(Y = b) = \sum_{b \in \mathcal{B}} f(b)\mathbb{P}(Y = b).$$

# Conditional Expectation

Let $X$ and $Y$ be random variables defined on the same probability space, taking values in $\mathcal{A}$ and $\mathcal{B}$, respectively.

**Definition (Conditional Expectation):** For any $b \in \mathcal{B}$ with $\mathbb{P}[Y = b] > 0$, the conditional expectation of $X$ given $Y = b$ is

$$\mathbb{E}[X \mid Y = b] = \sum_{a \in \mathcal{A}} a \ \mathbb{P}(X = a \mid Y = b).$$

Define $f(Y) = \mathbb{E}[X \mid Y]$, which takes value $\mathbb{E}[X \mid Y = b]$ when $Y = b$. Then,

$$\mathbb{E}[f(Y)] = \sum_{b \in \mathcal{B}} f(b)\mathbb{P}[Y = b] = \underbrace{\sum_{b \in \mathcal{B}} \mathbb{E}[X \mid Y = b]\mathbb{P}[Y = b]}_{\mathbb{E}[X]}$$

# Law of Total Expectation

**Theorem (Law of Total Expectation):** If $\mathbb{E}[|X|] < \infty$,

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X].$$

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[X \mid Y\right]\right] &= \sum_{b \in \mathcal{B}} \mathbb{E}\left[X \mid Y = b\right] \mathbb{P}[Y = b] \\
&= \sum_{b \in \mathcal{B}} \sum_{a \in \mathcal{A}} a \, \mathbb{P}[X = a \mid Y = b] \, \mathbb{P}[Y = b] \\
&= \sum_{a \in \mathcal{A}} a \sum_{b \in \mathcal{B}} \mathbb{P}[X = a \mid Y = b] \, \mathbb{P}[Y = b] \\
&= \sum_{a \in \mathcal{A}} a \, \mathbb{P}[X = a] \\
&= \mathbb{E}\left[X\right].
\end{aligned}$$

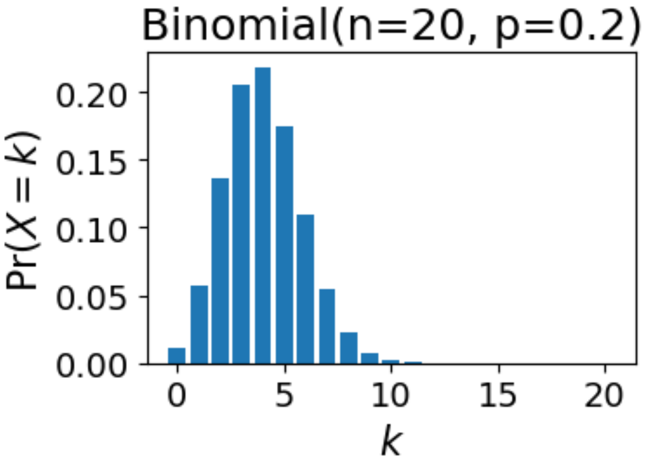
Interchange the order of summation

By the Total Probability Rule

Q: Why is this a useful result?

$\square$

# Variance




200 independent draws from Binomial(20, 0.2):

```
[7 4 4 4 3 8 5 6 3 3 6 3 4 2 4 7 2 4 3 2 3 2 3 3 4
2 5 5 1 5 1 4 4 4 6 6 4 4 2 8 5 7 7 3 6 3 6 4 3 3 3
4 2 5 5 5 1 4 3 3 2 4 2 3 4 3 3 4 1 3 2 4 3 2 3 7 7
7 9 6 6 4 8 5 6 2 2 2 5 4 3 3 7 4 4 3 5 3 2 4 4 2 6
1 6 3 5 3 6 3 3 1 3 1 5 3 3 3 6 6 7 5 1 5 2 3 3 5 4
4 4 6 7 5 4 5 3 1 4 2 2 6 6 3 2 1 2 8 1 4 2 1 3 5 3
4 2 1 2 7 7 2 5 5 2 1 4 1 5 6 4 3 2 5 2 5 5 3 2 6 5
7 6 2 3 5 4 8 5 7 3 4 4 4 5 6 1 6 7 4]
```

Sample average = 3.96, which is quite close to $\mathbb{E}[X] = np = 4$.

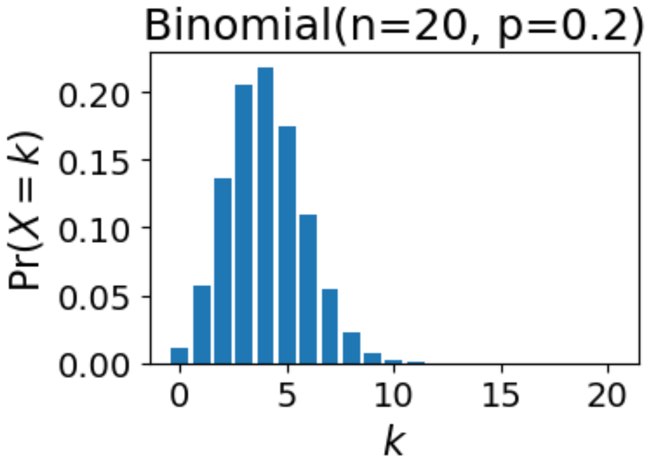
Next week, we will prove a precise relationship between sample averages and expectations.

## Signed deviations from $\mathbb{E}[X] = 4$:

```
[ 3, 0, 0, 0, −1, 4, 1, 2, −1, −1, 2, −1, 0, −2, 0, 3, −2, 0, −1, −2, −1, −2, −1, −1, 0, −2, 1, 1,
−3, 1, −3, 0, 0, 0, 2, 2, 0, 0, −2, 4, 1, 3, 3, −1, 2, −1, 2, 0, −1, −1, −1, 0, −2, 1, 1, 1, −3, 0,
−1, −1, −2, 0, −2, −1, 0, −1, −1, 0, −3, −1, −2, 0, −1, −2, −1, 3, 3, 3, 5, 2, 2, 0, 4, 1, 2, −2,
−2, −2, 1, 0, −1, −1, 3, 0, 0, −1, 1, −1, −2, 0, 0, −2, 2, −3, 2, −1, 1, −1, 2, −1, −1, −3, −1, −3,
1, −1, −1, −1, 2, 2, 3, 1, −3, 1, −2, −1, −1, 1, 0, 0, 0, 2, 3, 1, 0, 1, −1, −3, 0, −2, −2, 2, 2,
−1, −2, −3, −2, 4, −3, 0, −2, −3, −1, 1, −1, 0, −2, −3, −2, 3, 3, −2, 1, 1, −2, −3, 0, −3, 1, 2, 0,
−1, −2, 1, −2, 1, 1, −1, −2, 2, 1, 3, 2, −2, −1, 1, 0, 4, 1, 3, −1, 0, 0, 0, 1, 2, −3, 2, 3, 0]
```

$\boxed{\text{Average signed deviations} = -0.04}$ ← Will tend to zero as the sample size approaches $\infty$

# Variance




200 independent draws from Binomial(20, 0.2):

```
[7 4 4 4 3 8 5 6 3 3 6 3 4 2 4 7 2 4 3 2 3 2 3 3 4
2 5 5 1 5 1 4 4 4 6 6 4 4 2 8 5 7 7 3 6 3 6 4 3 3 3
4 2 5 5 5 1 4 3 3 2 4 2 3 4 3 3 4 1 3 2 4 3 2 3 7 7
7 9 6 6 4 8 5 6 2 2 2 5 4 3 3 7 4 4 3 5 3 2 4 4 2 6
1 6 3 5 3 6 3 3 1 3 1 5 3 3 3 6 6 7 5 1 5 2 3 3 5 4
4 4 6 7 5 4 5 3 1 4 2 2 6 6 3 2 1 2 8 1 4 2 1 3 5 3
4 2 1 2 7 7 2 5 5 2 1 4 1 5 6 4 3 2 5 2 5 5 3 2 6 5
7 6 2 3 5 4 8 5 7 3 4 4 4 5 6 1 6 7 4]
```

Sample average = 3.96, which is quite close to $\mathbb{E}[X] = np = 4$.

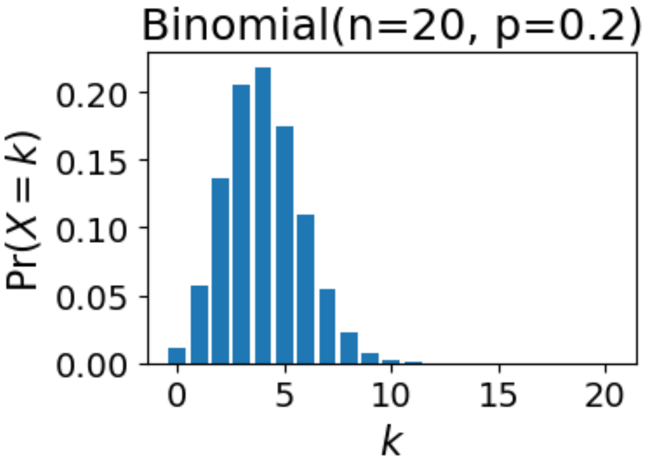
Next week, we will prove a precise relationship between sample averages and expectations

Squared deviations from $\mathbb{E}[X] = 4$:

```
[ 9, 0, 0, 0, 1, 16, 1, 4, 1, 1, 4, 1, 0, 4, 0, 9, 4, 0, 1, 4, 1, 4, 1, 1, 0, 4, 1, 1, 9, 1, 9,
0, 0, 0, 4, 4, 0, 0, 4, 16, 1, 9, 9, 1, 4, 1, 4, 0, 1, 1, 1, 0, 4, 1, 1, 1, 9, 0, 1, 1, 4, 0, 4,
1, 0, 1, 1, 0, 9, 1, 4, 0, 1, 4, 1, 9, 9, 9, 25, 4, 4, 0, 16, 1, 4, 4, 4, 4, 1, 0, 1, 1, 9, 0, 0,
1, 1, 1, 4, 0, 0, 4, 4, 9, 4, 1, 1, 1, 4, 1, 1, 9, 1, 9, 1, 1, 1, 1, 4, 4, 9, 1, 9, 1, 4, 1, 1,
1, 0, 0, 0, 4, 9, 1, 0, 1, 1, 9, 0, 4, 4, 4, 4, 1, 4, 9, 4, 16, 9, 0, 4, 9, 1, 1, 1, 0, 4, 9, 4,
9, 9, 4, 1, 1, 4, 9, 0, 9, 1, 4, 0, 1, 4, 1, 4, 1, 1, 1, 4, 4, 1, 9, 4, 4, 1, 1, 0, 16, 1, 9, 1,
0, 0, 0, 1, 4, 9, 4, 9, 0]
```

Average squared deviations = 3.32

# Variance




200 independent draws from Binomial(20, 0.2):

```
[7 4 4 4 3 8 5 6 3 3 6 3 4 2 4 7 2 4 3 2 3 2 3 3 4
 2 5 5 1 5 1 4 4 4 6 6 4 4 2 8 5 7 7 3 6 3 6 4 3 3 3
 4 2 5 5 5 1 4 3 3 2 4 2 3 4 3 3 4 1 3 2 4 3 2 3 7 7
 7 9 6 6 4 8 5 6 2 2 2 5 4 3 3 7 4 4 3 5 3 2 4 4 2 6
 1 6 3 5 3 6 3 3 1 3 1 5 3 3 3 6 6 7 5 1 5 2 3 3 5 4
 4 4 6 7 5 4 5 3 1 4 2 2 6 6 3 2 1 2 8 1 4 2 1 3 5 3
 4 2 1 2 7 7 2 5 5 2 1 4 1 5 6 4 3 2 5 2 5 5 3 2 6 5
 7 6 2 3 5 4 8 5 7 3 4 4 4 5 6 1 6 7 4]
```

Sample average = 3.96, which is quite close to $\mathbb{E}[X] = np = 4$.

Next week, we will prove a precise relationship between sample average and $\mathbb{E}[X]$.

As the sample size increases, the average squared deviations will approach:

**Definition (Variance):** Suppose $X$ is a random variable with $\mathbb{E}[X] = \mu$. Then, $\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2] \geq 0$. (Recall $\mathbb{E}[f(X)]$ and let $f(X) = (X-\mu)^2$.)

The standard deviation of $X$ is defined as $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.

# Variance

**Definition (Variance):** Suppose $X$ is a random variable with $\mathbb{E}[X] = \mu$. Then, $\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2] \geq 0$.

The standard deviation of $X$ is defined as $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.

Note: $\mathrm{Var}(X) = \mathbb{E}[X^2 - 2\mu X + \mu^2] \underset{\text{by linearity of expectation}}{=} \mathbb{E}[X^2] - 2\mu\underbrace{\mathbb{E}[X]}_{\mu} + \mu^2 = \mathbb{E}[X^2] - \mu^2$.

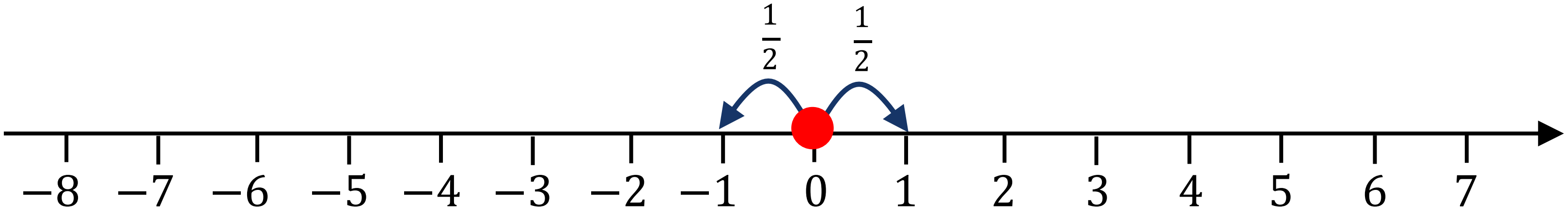
For given constants $a, b$ and a random variable $X$, let $Y = aX + b$.
What is $\mathrm{Var}(Y)$?

$$\mathbb{E}[Y] = \mathbb{E}[aX + b] = a\mathbb{E}[X] + b = a\mu + b$$

$$\mathrm{Var}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathbb{E}[(aX + b - a\mu + b)^2] = \mathbb{E}[a^2(X-\mu)^2]$$

by linearity of expectation $= a^2\mathbb{E}[(X-\mu)^2] = a^2\mathrm{Var}(X)$

# Symmetric Random Walk

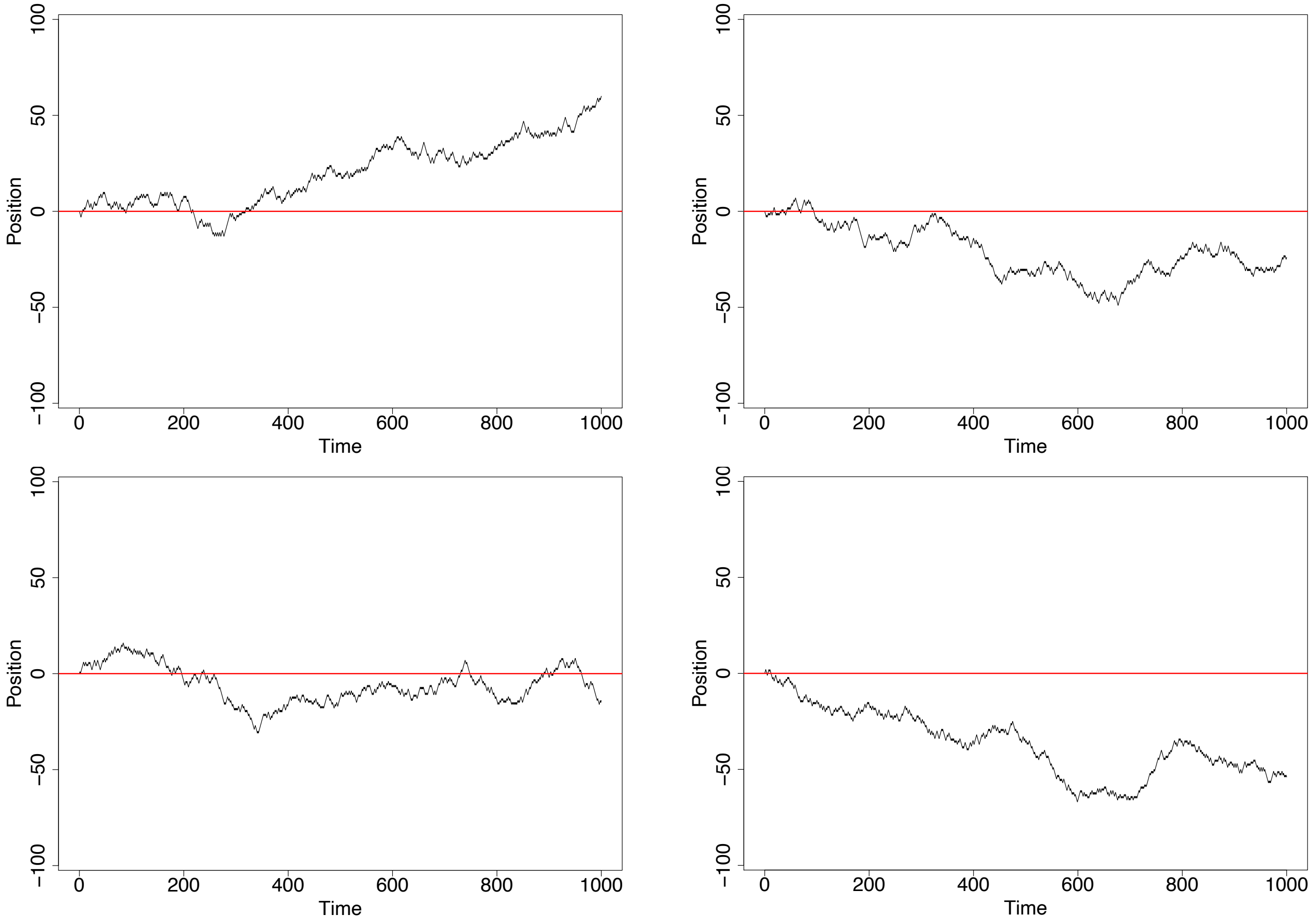



- Start at the origin.
- For each time step, walk to the left with probability ½ or to the right with probability ½, independently of all previous steps.
- Position after $n$ steps: $S_n = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ are i.i.d. random variables with

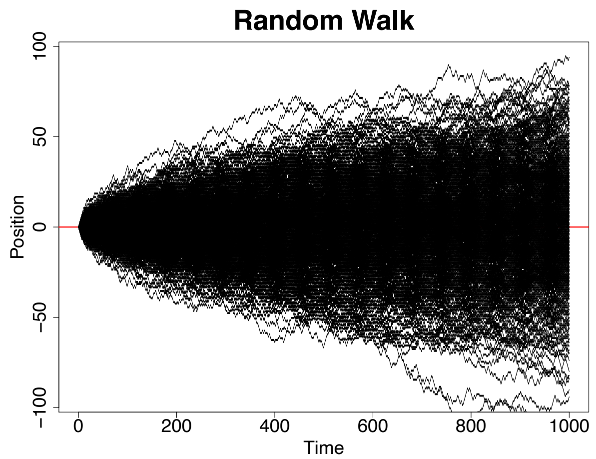
$$X_i = \begin{cases} +1, & \text{with probability } 1/2 \\ -1, & \text{with probability } 1/2 \end{cases}$$

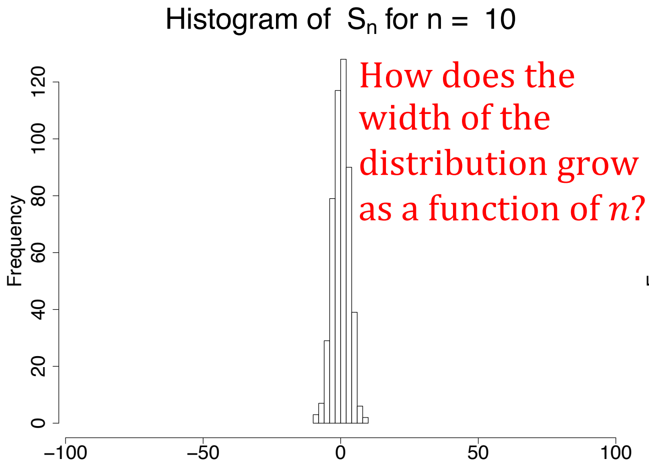
- What is $\mathbb{E}[S_n]$?
- How about $\mathrm{Var}[S_n]$?

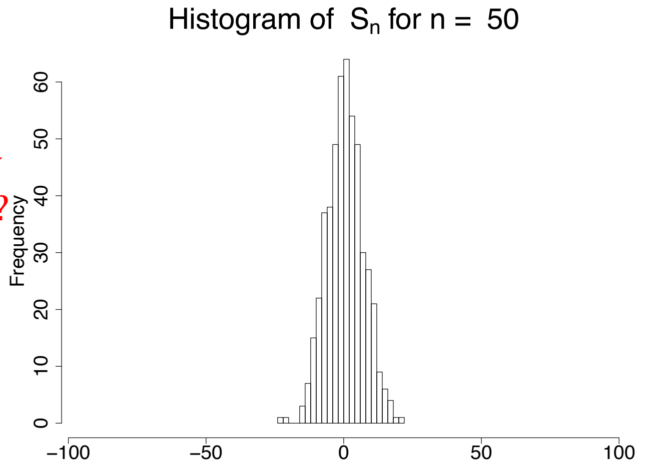
# Random Walk (Simulations)

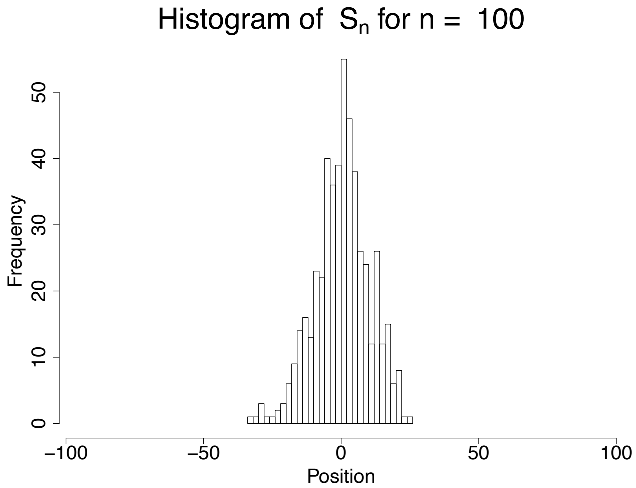


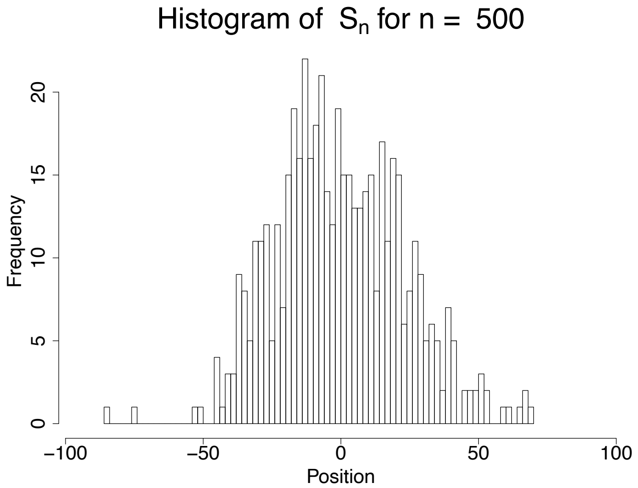
# Random Walk (Distributions of $S_n$)

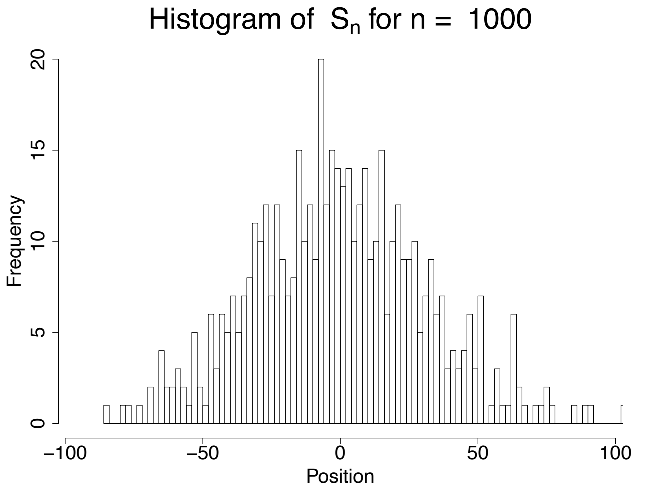


# Expectation of $XY$

**Lemma 1:** If $X$ and $Y$ are independent random variables on the same probability space, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

**Proof:** 
$$\mathbb{E}[XY] = \sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} ab\,\mathbb{P}(X=a, Y=b)$$
$$= \sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} ab\,\mathbb{P}(X=a)\mathbb{P}(Y=b) = \left[\sum_{a\in\mathcal{A}} a\mathbb{P}(X=a)\right]\left[\sum_{b\in\mathcal{B}} b\,\mathbb{P}(Y=b)\right] = \mathbb{E}[X]\mathbb{E}[Y]. \quad \square$$

---

Is the converse true? No!

| $(a,b)$ | $\mathbb{P}(X=a, Y=b)$ |
|---|---|
| $(-1,0)$ | $1/3$ |
| $(0,1)$ | $1/3$ |
| $(1,0)$ | $1/3$ |

| $a$ | $\mathbb{P}(X=a)$ |
|---|---|
| $-1$ | $1/3$ |
| $0$ | $1/3$ |
| $1$ | $1/3$ |

| $b$ | $\mathbb{P}(Y=b)$ |
|---|---|
| $0$ | $2/3$ |
| $1$ | $1/3$ |

However,
$$\left.\begin{aligned}\mathbb{E}[XY] &= 0\\ \mathbb{E}[X] &= 0\\ \mathbb{E}[Y] &= 1/3\end{aligned}\right\} \Rightarrow \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Are $X$ and $Y$ independent? No! $\underbrace{\mathbb{P}(X=1, Y=1)}^{0} \neq \overbrace{\mathbb{P}(X=1)}^{1/3}\overbrace{\mathbb{P}(Y=1)}^{1/3}$

# Variance of $X + Y$

**Theorem:** If $X$ and $Y$ are independent random variables on the same probability space, then $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

**Proof.** Using **linearity of expectation**, we get

$$\mathbb{E}[(X+Y)^2] - \left[\mathbb{E}[X+Y]\right]^2 = \mathbb{E}[X^2 + 2XY + Y^2] - \left[\mathbb{E}[X] + \mathbb{E}[Y]\right]^2$$

$$= \underbrace{\mathbb{E}[X^2] - \left[\mathbb{E}[X]\right]^2}_{\mathrm{Var}(X)} + \underbrace{\mathbb{E}[Y^2] - \left[\mathbb{E}[Y]\right]^2}_{\mathrm{Var}(Y)} + 2\underbrace{\left[\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\right]}_{\text{by Lemma 1}}$$

$\square$

---

Back to Symmetric Random Walk:

$S_n = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ are i.i.d. random variables with

$$X_i = \begin{cases} +1, & \text{with probability } 1/2 \\ -1, & \text{with probability } 1/2 \end{cases}$$

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = n \quad \Rightarrow \quad \sigma(S_n) = \sqrt{n}.$$

$$\mathrm{Var}(X_i) = \mathbb{E}\left[X_i^2\right] - (\mathbb{E}[X_i])^2 = (1)^2\mathbb{P}(X_i = 1) + (-1)^2\mathbb{P}(X_i = -1) = 1$$

# Variance Examples

**Example (Bernoulli):** $I \sim \text{Bernoulli}(p)$, for $0 < p < 1$.

$$\mathbb{P}(I = a) = \begin{cases} p, & \text{if } a = 1, \\ 1 - p, & \text{if } a = 0. \end{cases}$$

$\mathbb{E}[I] = p$

$\text{Var}(I) = \mathbb{E}[I^2] - (\mathbb{E}[I])^2 = p - p^2 = p(1 - p)$

**Example (Binomial):** $X \sim \text{Binomial}(n, p)$, where $n \in \mathbb{Z}_+$ and $0 < p < 1$.

For $k = 0, \ldots, n$, $\quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$.

$X = I_1 + \cdots + I_n$, where $I_1, \ldots, I_n$ are mutually independent Bernoulli$(p)$ random variables.

$$\text{Var}(X) = \sum_{k=1}^{n} \text{Var}(I_k) = np(1 - p).$$

# Covariance and Correlation

**Definition (Covariance):** The covariance of random variables $X$ and $Y$ is defined as

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

**Definition (Correlation):** The correlation of random variables $X$ and $Y$ with $\sigma(X) > 0$ and $\sigma(Y) > 0$ is defined as

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

**Theorem:** For a pair of random variables $X, Y$ with $\sigma(X) > 0$ and $\sigma(Y) > 0$,

$$-1 \leq \mathrm{Corr}(X,Y) \leq 1.$$

See Note 17 for proof.